US006243667B1

(12) **United States Patent**
Kerr et al.

(10) **Patent No.: US 6,243,667 B1**
(45) **Date of Patent: *Jun. 5, 2001**

(54) **NETWORK FLOW SWITCHING AND FLOW DATA EXPORT**

(75) Inventors: **Darren R. Kerr**, Union City; **Barry L. Bruins**, Los Altos, both of CA (US)

(73) Assignee: **Cisco Systems, Inc.**, San Jose, CA (US)

(*) Notice: This patent issued on a continued prosecution application filed under 37 CFR 1.53(d), and is subject to the twenty year patent term provisions of 35 U.S.C. 154(a)(2).

Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **08/655,429**

(22) Filed: **May 28, 1996**

(51) **Int. Cl.**[7] **G06F 9/34**

(52) **U.S. Cl.** **703/27**; 703/20; 370/379; 370/392; 370/389

(58) **Field of Search** 395/500, 200.01, 395/200.13, 683, 185.04; 370/352, 389, 392, 351, 410

(56) **References Cited**

U.S. PATENT DOCUMENTS

Re. 33,900	4/1992	Howson	370/105
4,131,767	12/1978	Weinstein	179/170.2
4,161,719	7/1979	Parikh et al.	340/147 SY

(List continued on next page.)

FOREIGN PATENT DOCUMENTS

0 384 758	2/1990	(EP)	H04L/12/56
0 431 751 A1	11/1990	(EP)	H04I/12/46
WO 95/20850	8/1995	(WO)	H04L/12/56

OTHER PUBLICATIONS

Cormen et al., "Introduction to Algorithms", MIT Press, seventeenth edition, pp. 221–224.*

Pei et al., VLSI Implementation of Routing Tables: Tries and Cams, IEEE, 1991, pp. 515–524.*

Chandranmenon et al., "Trading Packet Headers for Packet Processing," IEEE, Apr. 1996, pp. 141–152.*

Cao et al., Performance of Hashing–Based Schemes for Internet Load Balancing, IEEE, 2000, pp. 332–341.*

Newman et al., "Flow Labelled IP: A Connectionless Approach to ATM," IEEE, Mar. 1996, pp. 1251–1260.*

Newman et al., "IP Switching and Gigabit Routers," IEEE, 1997, pp. 64–69.*

Worster et al., "Levels of Aggregation in Flow Switching Networks," IEEE, 1997, pp. 51–59.*

William Stallings, Data and Computer Communications, pp. 329–333, Prentice Hall, Upper Saddle River, New Jersey 07458.

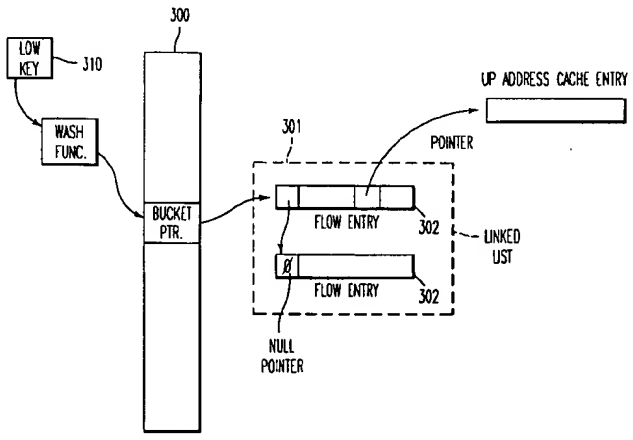
(List continued on next page.)

Primary Examiner—Kevin J. Teska
Assistant Examiner—Thai Phan
(74) *Attorney, Agent, or Firm*—Oblon, Spivak, McClelland, Maier & Neustadt, P.C.

(57) **ABSTRACT**

The invention provides a method and system for switching in networks responsive to message flow patterns. A message "flow" is defined to comprise a set of packets to be transmitted between a particular source and a particular destination. When routers in a network identify a new message flow, they determine the proper processing for packets in that message flow and cache that information for that message flow. Thereafter, when routers in a network identify a packet which is part of that message flow, they process that packet according to the proper processing for packets in that message flow. The proper processing may include a determination of a destination port for routing those packets and a determination of whether access control permits routing those packets to their indicated destination.

19 Claims, 5 Drawing Sheets



U.S. PATENT DOCUMENTS

4,316,284	2/1982	Howson	370/105
4,397,020	8/1983	Howson	370/105
4,419,728	12/1983	Larson	364/200
4,424,565	1/1984	Larson	364/200
4,437,087	3/1984	Petr	340/347 DD
4,438,511	3/1984	Baran	370/19
4,439,763	3/1984	Limb	340/825.5
4,445,213	4/1984	Baugh et al.	370/94
4,446,555	5/1984	Devault et al.	370/94
4,456,957	6/1984	Schieltz	364/200
4,464,658	8/1984	Thelen	340/825.5
4,499,576	2/1985	Fraser	370/60
4,506,358	3/1985	Montgomery	370/60
4,507,760	3/1985	Fraser	365/221
4,532,626	7/1985	Flores et al.	370/85
4,644,532	2/1987	George et al.	370/94
4,646,287	2/1987	Larson et al.	370/60
4,677,423	6/1987	Benvenuto et al.	340/347 DD
4,679,189 *	7/1987	Olson et al.	370/60
4,679,227	7/1987	Hughes-Hartogs	379/98
4,723,267	2/1988	Jones et al.	379/93
4,731,816	3/1988	Hughes-Hartogs	379/98
4,750,136	6/1988	Arpin et al.	364/514
4,757,495	7/1988	Decker et al.	370/76
4,763,191	8/1988	Gordon et al.	358/86
4,769,810	9/1988	Eckberg, Jr. et al.	370/60
4,769,811	9/1988	Eckberg, Jr. et al.	370/60
4,771,425	9/1988	Baran et al.	370/85
4,819,228	4/1989	Baran et al.	370/85
4,827,411	5/1989	Arrowhead et al.	364/300
4,833,706	5/1989	Hughes-Hartogs	379/98
4,835,737	5/1989	Herrig et al.	364/900
4,879,551	11/1989	Georgiou et al.	340/825.87
4,893,306	1/1990	Chao et al.	340/94.2
4,903,261	2/1990	Baran et al.	370/94.2
4,922,486	5/1990	Lidinsky et al.	370/60
4,933,937	6/1990	Konishi	370/85.13
4,960,310	10/1990	Cushing	350/1.7
4,962,497	10/1990	Ferenc et al.	370/60.1
4,962,532	10/1990	Kasiraj et al.	380/25
4,965,767	10/1990	Kinoshita et al. .	
4,965,772	10/1990	Daniel et al.	364/900
4,970,678	11/1990	Sladowski et al.	364/900
4,979,118 *	12/1990	Kheradpir	364/436
4,980,897	12/1990	Decker et al.	375/38
4,991,169	2/1991	Davis et al.	370/77
5,003,595	3/1991	Collins et al.	380/25
5,014,265	5/1991	Hahne et al.	370/60
5,020,058	5/1991	Holden et al.	370/109
5,033,076	7/1991	Jones et al.	379/67
5,034,919	7/1991	Sasai et al. .	
5,054,034	10/1991	Hughes-Hartogs	375/8
5,059,925	10/1991	Weisbloom	331/1 A
5,072,449	12/1991	Enns et al.	371/37.1
5,088,032	2/1992	Bosack	395/200
5,095,480 *	3/1992	Fenner et al.	370/94.1
5,115,431	5/1992	Williams et al.	370/94.1
5,128,945	7/1992	Enns et al.	371/37.1
5,136,580	8/1992	Videlock et al.	370/60
5,166,930	11/1992	Braff et al.	370/94.1
5,199,049	3/1993	Wilson	375/104
5,206,886	4/1993	Bingham	375/97
5,208,811	5/1993	Kashio et al. .	
5,212,686	5/1993	Joy et al.	370/60
5,224,099	6/1993	Corbalis et al.	370/94.2
5,226,120	7/1993	Brown et al.	395/200
5,228,062	7/1993	Bingham	375/97
5,229,994	7/1993	Balzano et al.	370/85.13
5,237,564	8/1993	Lespagnol et al.	370/60.1
5,241,682	8/1993	Bryant et al.	395/800
5,243,342	9/1993	Kattemalalavadi et al.	341/106
5,243,596	9/1993	Port et al.	370/94.1
5,247,516	9/1993	Bernstein et al.	370/82
5,249,178	9/1993	Kurano et al.	370/60
5,249,292 *	9/1993	Chiappa	395/650
5,253,251	10/1993	Aramaki .	
5,255,291	10/1993	Holden et al.	375/111
5,260,933	11/1993	Rouse	370/14
5,260,978	11/1993	Fleischer et al.	375/106
5,268,592	12/1993	Bellamy et al.	307/43
5,268,900	12/1993	Hluchyj et al.	370/94.1
5,271,004	12/1993	Proctor et al.	370/60
5,274,631	12/1993	Bhardwaj	370/60
5,274,635	12/1993	Rahman et al.	370/60.1
5,274,643	12/1993	Fisk	370/94.1
5,280,470	1/1994	Buhrke et al.	370/13
5,280,480	1/1994	Pitt et al.	370/85.13
5,280,500	1/1994	Mazzola et al.	375/17
5,283,783	2/1994	Nguyen et al.	370/16.1
5,287,103	2/1994	Kasprzyk et al.	340/825.52
5,287,453 *	2/1994	Roberts	395/200
5,291,482	3/1994	McHarg et al.	370/60
5,305,311	4/1994	Lyles	370/60
5,307,343	4/1994	Bostica et al.	370/60
5,309,437 *	5/1994	Perlman et al.	370/85.13
5,311,509	5/1994	Heddes et al.	370/60
5,313,454	5/1994	Bustini et al.	370/13
5,313,582	5/1994	Hendel et al.	395/250
5,317,562	5/1994	Nardin et al.	370/16
5,319,644	6/1994	Liang	370/85.5
5,327,421	7/1994	Hiller et al.	370/60.1
5,331,637	7/1994	Francis et al. .	
5,345,445	9/1994	Hiller et al.	370/60.1
5,345,446	9/1994	Hiller et al.	370/60.1
5,359,592	10/1994	Corbalis et al.	370/17
5,361,250	11/1994	Nguyen et al.	370/16.1
5,361,256	11/1994	Doeringer et al. .	
5,361,259	11/1994	Hunt et al.	370/84
5,365,524	11/1994	Hiller et al.	370/94.2
5,367,517	11/1994	Cidon et al.	370/54
5,371,852	12/1994	Attanasio et al.	395/200
5,386,567	1/1995	Lien et al.	395/700
5,390,170	2/1995	Sawant et al.	370/58.1
5,390,175	2/1995	Hiller et al.	370/60
5,394,394	2/1995	Crowther et al.	370/60
5,394,402	2/1995	Ross	370/94.1
5,400,325	3/1995	Chatwani et al.	370/60.1
5,408,469	4/1995	Opher et al.	370/60.1
5,416,842	5/1995	Aziz	380/30
5,422,880	6/1995	Heitkamp et al.	370/60
5,422,882	6/1995	Hiller et al.	370/60.1
5,423,002	6/1995	Hart	395/200
5,426,636	6/1995	Hiller et al.	370/60.1
5,426,637 *	6/1995	Derby et al.	370/85.13
5,428,607	6/1995	Hiller et al.	370/60.1
5,430,715	7/1995	Corbalis et al.	370/54
5,430,729	7/1995	Rahnema .	
5,442,457	8/1995	Najafi	385/400
5,442,630	8/1995	Gagliardi et al.	370/85.13
5,452,297	9/1995	Hiller et al.	370/60.1
5,473,599	12/1995	Li et al.	370/16
5,473,607	12/1995	Hausman et al.	370/85.13
5,477,541	12/1995	White et al. .	
5,485,455 *	1/1996	Dobbins et al.	370/60
5,490,140	2/1996	Abensour et al. .	
5,490,258 *	2/1996	Fenner	395/401
5,491,687	2/1996	Christensen et al.	370/17
5,491,693 *	2/1996	Britton et al.	370/85.13
5,491,804	2/1996	Heath et al.	395/275
5,497,368	3/1996	Reijnierse et al. .	
5,504,747	4/1996	Sweazey .	

5,509,006		4/1996	Wilford et al.	370/60
5,509,123	*	4/1996	Dobbins et al.	395/200.15
5,517,494		5/1996	Green .	
5,517,662	*	5/1996	Coleman et al.	395/800
5,519,704		5/1996	Farinacci et al.	370/85.13
5,519,858	*	5/1996	Walton et al.	395/600
5,524,254	*	6/1996	Morgan et al.	395/800
5,526,489		6/1996	Nilakantan et al.	395/200.02
5,530,963		6/1996	Moore et al.	395/200.15
5,535,195		7/1996	Lee	370/54
5,539,734		7/1996	Burwell et al. .	
5,541,911		7/1996	Nilakantan et al. .	
5,546,370		8/1996	Ishikawa .	
5,550,816	*	8/1996	Hardwick et al.	370/60
5,555,244		9/1996	Gupta et al.	370/60
5,561,669		10/1996	Lenney et al.	370/60.1
5,583,862		12/1996	Callon	370/397
5,592,470		1/1997	Rudrapatna et al.	370/320
5,598,581		1/1997	Daines et al.	395/872
5,600,798		2/1997	Chenrukuri et al. .	
5,602,770		2/1997	Ohira .	
5,604,868		2/1997	Komine et al.	395/200
5,608,726		3/1997	Virgile .	
5,617,417		4/1997	Sathe et al.	370/394
5,617,421		4/1997	Chin et al.	370/402
5,630,125	*	5/1997	Zellweger	395/614
5,631,908		5/1997	Saxe .	
5,632,021		5/1997	Jennings et al.	395/309
5,634,010		5/1997	Ciscon et al.	395/200
5,634,011	*	5/1997	Auerbach et al.	395/200.15
5,644,718		7/1997	Belove et al.	395/200
5,666,353		9/1997	Klausmeier et al.	370/230
5,673,265		9/1997	Gupta et al.	370/432
5,675,579	*	10/1997	Watson et al.	370/248
5,678,006		10/1997	Valizadeh et al.	395/200
5,680,116		10/1997	Hashimoto et al. .	
5,684,800	*	11/1997	Dobbins et al.	370/401
5,687,324		11/1997	Green et al. .	
5,724,351		3/1998	Chao et al. .	
5,740,097		4/1998	Satoh .	
5,748,186	*	5/1998	Raman	345/302
5,754,547		5/1998	Nakazawa .	
5,802,054		9/1998	Bellenger .	
5,835,710	*	11/1998	Nagami et al.	395/200.8
5,841,874		11/1998	Kempke et al. .	
5,854,903		12/1998	Morrison et al. .	
5,856,981		1/1999	Voelker .	
5,892,924	*	4/1999	Lyon et al.	395/200.75
5,898,686		4/1999	Virgile .	
5,903,559		5/1999	Acharya et al. .	
5,925,097	*	7/1999	Gopinath et al.	709/200

OTHER PUBLICATIONS

Chowdhury, et al., "Alternative Banddwidth Allocation Algorithms for Packet Video in ATM Networks", 1992, IEEE Infocom 92, pp. 1061–1068.

Zhang, et al., "Rate–Controlled Static–Priority Queueing", 1993, IEEE, pp. 227–236.

Doeringer, et al., "Routing on Longest–Matching Prefixes", IEEE ACM Transactions on Networking, Feb. 1, 1996, vol. 4, No. 1, pp. 86–97.

IBM, "Method and Apparatus for the Statistical Multiplexing of Voice, Data, and Image Signals", Nov., 1992, IBM Technical Data Bulletin n6, pp. 409–411.

Esaki, et al., "Datagram Delivery in an ATM–Internet," IEICE Transactions on Communications vol. E77–B, No. 3, (1994) Mar., Tokyo, Japan.

* cited by examiner

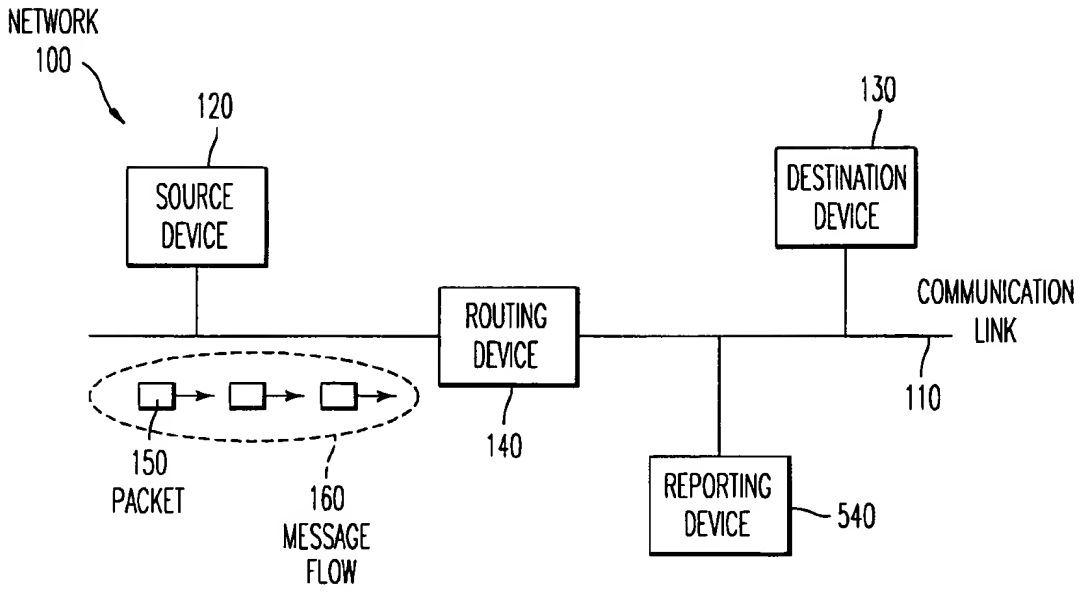

NETWORK
100
120
SOURCE DEVICE
130
DESTINATION DEVICE
ROUTING DEVICE
140
COMMUNICATION LINK
110
150
PACKET
160
MESSAGE FLOW
REPORTING DEVICE
540

FIG. 1

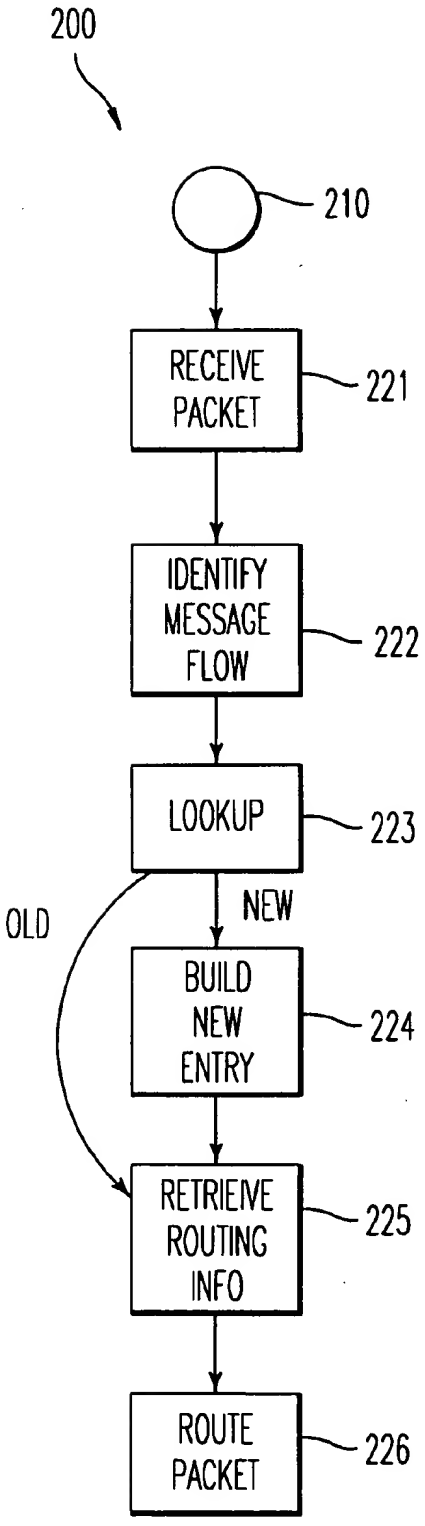

200
210
RECEIVE PACKET
221
IDENTIFY MESSAGE FLOW
222
LOOKUP
223
OLD
NEW
BUILD NEW ENTRY
224
RETRIEVE ROUTING INFO
225
ROUTE PACKET
226

FIG. 2A

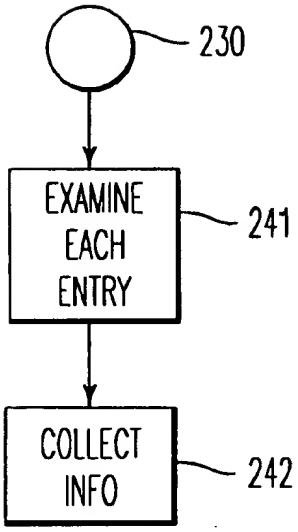

230
EXAMINE EACH ENTRY
241
COLLECT INFO
242

FIG. 2B

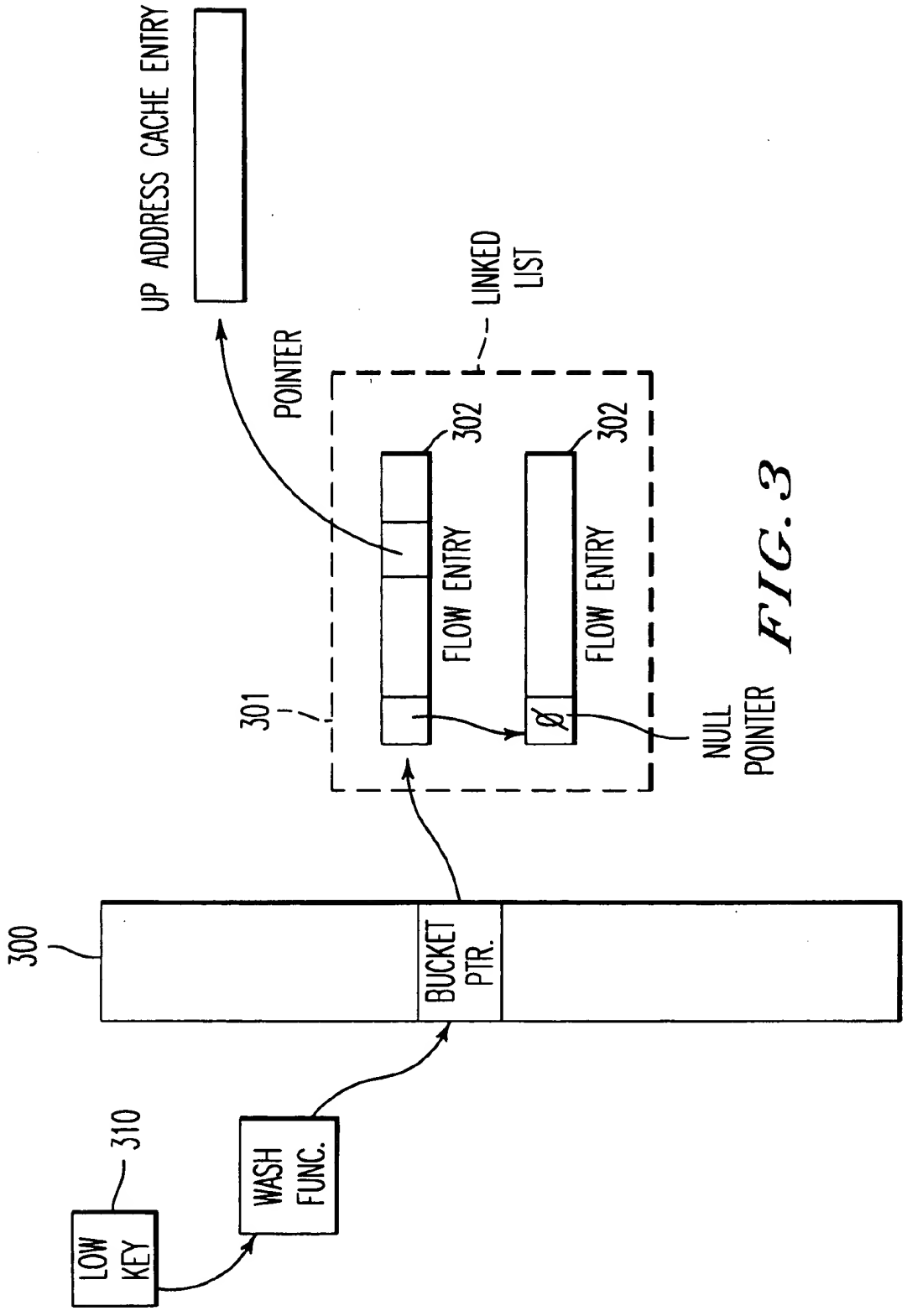

UP ADDRESS CACHE ENTRY
LINKED LIST
POINTER
302
302
FLOW ENTRY
FLOW ENTRY
301
NULL POINTER
300
BUCKET PTR.
310
WASH FUNC.
LOW KEY

FIG. 3

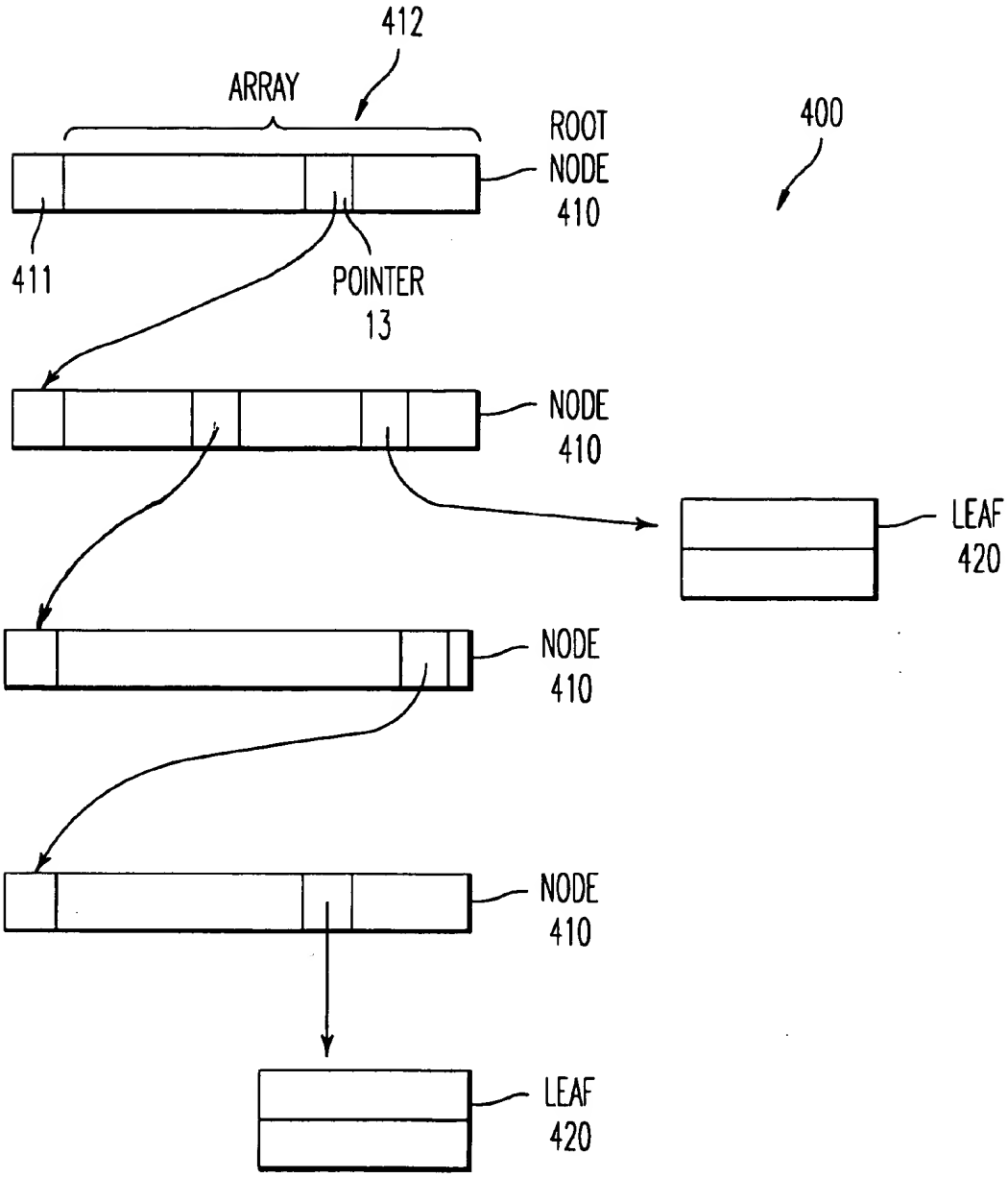

412
ARRAY
ROOT
NODE
410
400
411
POINTER
13
NODE
410
LEAF
420
NODE
410
NODE
410
LEAF
420

FIG. 4

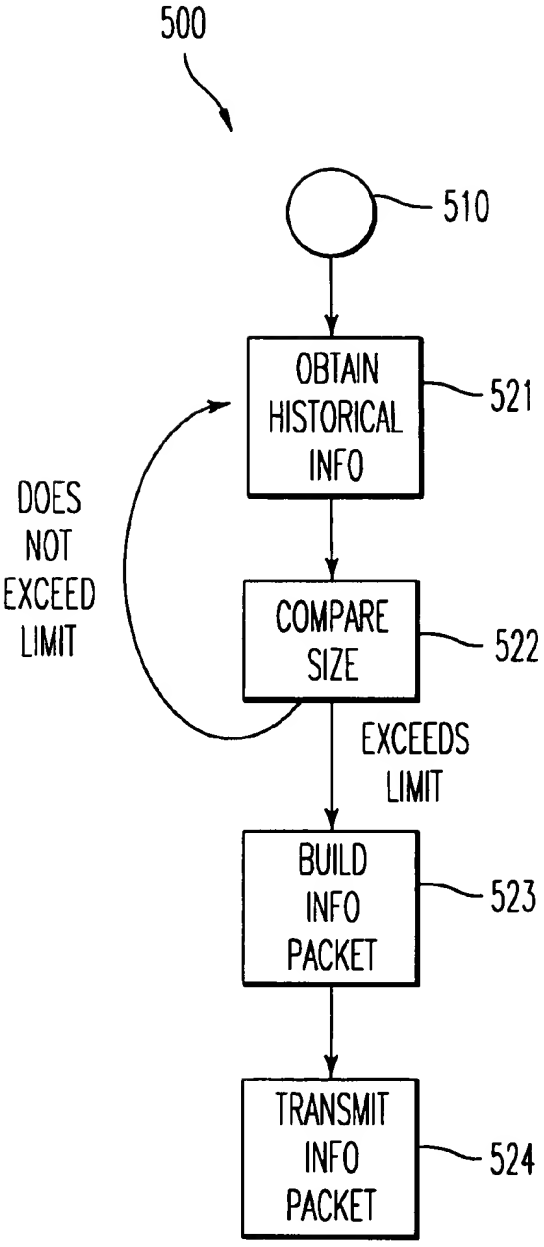

500
510
OBTAIN HISTORICAL INFO
521
DOES NOT EXCEED LIMIT
COMPARE SIZE
522
EXCEEDS LIMIT
BUILD INFO PACKET
523
TRANSMIT INFO PACKET
524

FIG. 5A

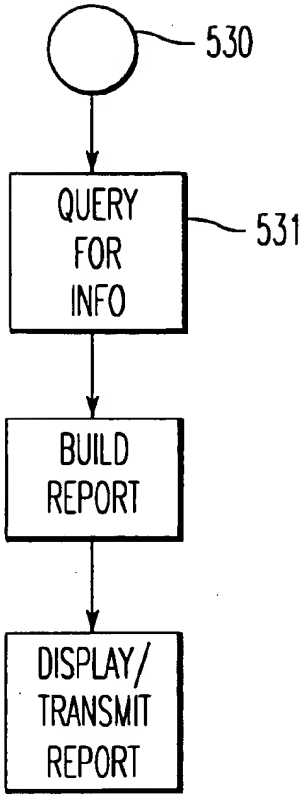

530
QUERY FOR INFO
531
BUILD REPORT
DISPLAY/ TRANSMIT REPORT

FIG. 5B

NETWORK FLOW SWITCHING AND FLOW DATA EXPORT

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates to network switching and data export responsive to message flow patterns.

2. Description of Related Art

In computer networks, it commonly occurs that message traffic between a particular source and a particular destination will continue for a time with unchanged routing or switching parameters. For example, when using the file-transfer protocol "FTP" there is substantial message traffic between the file's source location and the file's destination location, comprising the transfer of many packets which have similar headers, differing in the actual data which is transmitted. During the time when message traffic continues, routing and switching devices receiving packets comprising that message traffic must examine those packets and determine the processing thereof.

One problem which has arisen in the art is that processing demands on routing and switching devices continue to grow with increased network demand. It continues to be advantageous to provide techniques for processing packets more quickly. This problem has been exacerbated by addition of more complex forms of processing, such as the use of access control lists.

It would therefore be advantageous to provide techniques in which the amount of processing required for any individual packet could be reduced. With inventive techniques described herein, information about message flow patterns is used to identify packets for which processing has already been determined, and therefore to process those packets without having to re-determine the same processing. The amount of processing required for any individual packet is therefore reduced.

Information about message flow patterns would also be valuable for providing information about use of the network, and could be used for a variety of purposes by network administrators, routing devices, service providers, and users.

Accordingly, it would be advantageous to provide a technique for network switching and data export responsive to message flow patterns.

SUMMARY OF THE INVENTION

The invention provides a method and system for switching in networks responsive to message flow patterns. A message "flow" is defined to comprise a set of packets to be transmitted between a particular source and a particular destination. When routers in a network identify a new message flow, they determine the proper processing for packets in that message flow and cache that information for that message flow. Thereafter, when routers in a network identify a packet which is part of that message flow, they process that packet according to the proper processing for packets in that message flow. The proper processing may include a determination of a destination port for routing those packets and a determination of whether access control permits routing those packets to their indicated destination.

In another aspect of the invention, information about message flow patterns is collected, responsive to identified message flows and their packets. The collected information is reported to devices on the network. The collected information is used for a variety of purposes, including: to diagnose actual or potential network problems, to determine

patterns of usage by date and time or by location, to determine which services and which users use a relatively larger or smaller amount of network resources, to determine which services are accessed by particular users, to determine which users access particular services, or to determine usage which falls within selected parameters (such as: access during particular dates or times, access to prohibited services, excessive access to particular services, excessive use of network resources, or lack of proper access).

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** shows a network in which routing responsive to message flow patterns is performed.

FIG. **2** shows a method for routing in networks responsive to message flow patterns.

FIG. **3** shows data structures for use with a method for routing in networks responsive to message flow patterns.

FIG. **4** shows an IP address cache for use with a method for routing in networks responsive to message flow patterns.

FIG. **5** shows a method for collecting and reporting information about message flow patterns.

DESCRIPTION OF THE PREFERRED EMBODIMENT

In the following description, a preferred embodiment of the invention is described with regard to preferred process steps and data structures. However, those skilled in the art would recognize, after perusal of this application, that embodiments of the invention may be implemented using a set of general purpose computers operating under program control, and that modification of a set of general purpose computers to implement the process steps and data structures described herein would not require undue invention.

Message Flows

FIG. **1** shows a network in which routing responsive to message flow patterns is performed.

A network **100** includes at least one communication link **110**, at least one source device **120**, at least one destination device **130**, and at least one routing device **140**. The routing device **140** is disposed for receiving a set of packets **150** from the source device **120** and routing them to the destination device **130**.

The communication link **110** may comprise any form of physical media layer, such as ethernet, FDDI, or HDLC serial link.

The routing device **140** comprises a routing processor for performing the process steps described herein, and may include specific hardware constructed or programmed performing the process steps described herein, a general purpose processor operating under program control, or some combination thereof.

A message flow **160** consists of a unidirectional stream of packets **150** to be transmitted between particular pairs of transport service access points (thus, network-layer addresses and port numbers). In a broad sense, a message flow **160** thus refers to a communication "circuit" between communication end-points. In a preferred embodiment, a message flow **160** is defined by a network-layer address for a particular source device **120**, a particular port number at the source device **120**, a network-layer address for a particular destination device **130**, a particular port number at the destination device **130**, and a particular transmission protocol type. For example, the transmission protocol type may

identify a known transmission protocol, such as UDP, TCP, ICMP, or IGMP (internet group management protocol).

In a preferred embodiment for use with a network of networks (an "internet"), the particular source device **120** is identified by its IP (internet protocol) address. The particular port number at the source device **120** is identified by either a port number which is specific to a particular process, or by a standard port number for the particular transmission protocol type. For example, a standard port number for the TCP protocol type is **6** and a standard port number for the UDP protocol type is **17**. Other protocols which may have standard port numbers include the FTP protocol, the TELNET protocol, an internet telephone protocol, or an internet video protocol such as the "CUSeeMe" protocol; these protocols are known in the art of networking. Similarly, the particular destination device **130** is identified by its IP (internet protocol) address; the particular port number at the destination device **130** is identified by either a port number which is specific to a particular process, or a standard port number for the particular transmission protocol type.

It will be clear to those skilled in the art, after perusing this application, that the concept of a message flow is quite broad, and encompasses a wide variety of possible alternatives within the scope and spirit of the invention. For example, in alternative embodiments, a message flow may be bi-directional instead of unidirectional, a message flow may be identified at a different protocol layer level than that of transport service access points, or a message flow may be identified responsive to other factors. These other factors may include one or more of the following: information in packet headers, packet length, time of packet transmission, or routing conditions on the network (such as relative network congestion or administrative policies with regard to routing and transmission).

Network Flow Switching

FIG. **2** shows a method for routing in networks responsive to message flow patterns.

In broad overview, the method for routing in networks responsive to message flow patterns comprises two parts. In a first part, the routing device **140** builds and uses a flow cache described in further detail with regard to FIG. **3**), in which routing information to be used for packets **150** in each particular message flow **160** is recorded and from which such routing information is retrieved for use. In a second part, the routing device **140** maintains the flow cache, such as by removing entries for message flows **160** which are no longer considered valid.

A method **200** for routing in networks responsive to message flow patterns is performed by the routing device **140**.

At a flow point **210**, the routing device **140** is disposed for building and using the flow cache.

At a step **221**, the routing device **140** receives a packet **150**.

At a step **222**, the routing device **140** identifies a message flow **160** for the packet **150**. In a preferred embodiment, the routing device **140** examines a header for the packet **150** and identifies the IP address for the source device **120**, the IP address for the destination device **130**, and the protocol type for the packet **150**. The routing device **140** determines the port number for the source device **120** and the port number for the destination device **130** responsive to the protocol type. Responsive to this set of information, the routing device **140** determines a flow key **310** (described with reference to FIG. **3**) for the message flow **160**.

At a step **223**, the routing device **140** performs a lookup in a flow cache for the identified message flow **160**. If the lookup is unsuccessful, the identified message flow **160** is a "new" message flow **160**, and the routing device **140** continues with the step **224**. If the lookup is successful, the identified message flow **160** is an "old" message flow **160**, and the routing device **140** continues with the step **225**.

In a preferred embodiment, the routing device **140** determines a hash table key responsive to the flow key **310**. This aspect of the step **223** is described in further detail with regard to FIG. **3**.

At a step **224**, the routing device **140** builds a new entry in the flow cache. The routing device **140** determines proper treatment of packets **150** in the message flow **160** and enters information regarding such proper treatment in a data structure pointed to by the new entry in the flow cache. In a preferred embodiment, the routing device **140** determines the proper treatment by performing a lookup in an IP address cache as shown in FIG. **4**.

In a preferred embodiment, the proper treatment of packets **150** in the message flow **160** includes treatment with regard to switching (thus, the routing device **140** determines an output port for switching packets **150** in the message flow **160**), with regard to access control (thus, the routing device **140** determines whether packets **150** in the message flow **160** meet the requirements of access control, as defined by access control lists in force at the routing device **140**), with regard to accounting (thus, the routing device **140** creates an accounting record for the message flow **160**), with regard to encryption (thus, the routing device **140** determines encryption treatment for packets **150** in the message flow **160**), and any special treatment for packets **150** in the message flow **160**.

In a preferred embodiment, the routing device **140** performs any special processing for new message flows **160** at this time. For example, in one preferred embodiment, the routing device **140** requires that the source device **120** or the destination device **130** must authenticate the message flow **160**. In that case, the routing device **140** transmits one or more packets **150** to the source device **120** or the destination device **130** to request information (such as a user identifier and a password) to authenticate the new message flow **160**, and receives one or more packets **150** comprising the authentication information. This technique could be useful for implementing security "firewalls" and other authentication systems.

Thereafter, the routing device **140** proceeds with the step **225**, using the information from the new entry in the flow cache, just as if the identified message flow **160** were an "old" message flow **160** and the lookup in a flow cache had been successful.

At a step **225**, the routing device **140** retrieves routing information from the entry in the flow cache for the identified message flow **160**.

In a preferred embodiment, the entry in the flow cache includes a pointer to a rewrite function for at least part of a header for the packet **150**. If this pointer is non-null, the routing device **140** invokes the rewrite function to alter the header for the packet **150**.

At a step **226**, the routing device **140** routes the packet **150** responsive to the routing information retrieved at the step **225**.

Thus, in a preferred embodiment, the routing device **140** does not separately determine, for each packet **150** in the message flow **160**, the information stored in the entry in the flow cache. Rather, when routing a packet **150** in the

message flow **160**, the routing device **140** reads the information from the entry in the flow cache and treats the packet **150** according to the information in the entry in the flow cache.

Thus, in a preferred embodiment, the routing device **140** routes the packet **150** to an output port, determines whether access is allowed for the packet **150**, determines encryption treatment for the packet **150**, and performs any special treatment for the packet **150**, all responsive to information in the entry in the flow cache.

In a preferred embodiment, the routing device **140** also enters accounting information in the entry in the flow cache for the packet **150**. When routing each packet **150** in the message flow **160**, the routing device **140** records the cumulative number of packets **150** and the cumulative number of bytes for the message flow **160**.

Because the routing device **140** processes each packet **150** in the message flow **160** responsive to the entry for the message flow **160** in the flow cache, the routing device **140** is able to implement administrative policies which are designated for each message flow **160** rather than for each packet **150**. For example, the routing device **140** is able to reserve specific amounts of bandwidth for particular message flows **160** and to queue packets **150** for transmission responsive to the bandwidth reserved for their particular message flows **160**.

Because the routing device **140** is able to associate each packet **150** with a particular message flow **160** and to associate each message flow **160** with particular network-layer source and destination addresses, the routing device **140** is able to associate network usage with particular workstations (and therefore with particular users) or with particular services available on the network. This can be used for accounting purposes, for enforcing administrative policies, or for providing usage information to interested parties.

For a first example, the routing device **140** is able to monitor and provide usage information regarding access using the HTTP protocol to world wide web pages at particular sites.

For a second example, the routing device **140** is able to monitor usage information regarding relative use of network resources, and to give priority to those message flows **160** which use relatively fewer network resources. This can occur when a first message flow **160** is using a relatively low-bandwidth transmission channel (such as a **28.8** kilobits per second modem transmission channel) and when a second message flow **160** is using a relatively high-bandwidth transmission channel (such as a T-1 transmission line).

At a flow point **230**, the routing device **140** is disposed for maintaining the flow cache.

At a step **241**, the routing device **140** examines each entry in the flow cache and compares a current time with a last time a packet **150** was routed using that particular entry. If the difference exceeds a first selected timeout, the message flow **160** represented by that entry is considered to have expired due to nonuse and thus to no longer be valid.

In a preferred embodiment, the routing device **140** also examines the entry in the flow cache and compares a current time with a first time a packet **150** was routed using that particular entry. If the difference exceeds a second selected timeout, the message flow **160** represented by that entry is considered to have expired due to age and thus to no longer be valid. The second selected timeout is preferably about one minute.

Expiring message flows **160** due to age artificially requires that a new message flow **160** must be created for the next packet **150** in the same communication session represented by the old message flow **160** which was expired. However, it is considered preferable to do so because it allows information to be collected and reported about message flows **160** without having to wait for those message flows **160** to expire from nonuse. For example, a multiple-broadcast communication session could reasonably last well beyond the time message flows **160** are expired for age, and if not so expired would mean that information about network usage would not account for significant network usage.

In a preferred embodiment, the routing device **140** also examines the entry in the flow cache and determines if the "next hop" information has changed. If so, the message flow **160** is expired due to changed conditions. Other changed conditions which might cause a message flow **160** to be expired include changes in access control lists or other changes which might affect the proper treatment of packets **150** in the message flow **160**. The routing device **140** also expires entries in the flow cache on a least-recently-used basis if the flow cache becomes too full.

If the message flow **160** is still valid, the routing device **140** continues with the next entry in the flow cache until all entries have been examined. If the message flow **160** is no longer valid, the routing device **140** continues with the step **242**.

At a step **242**, the routing device **140** collects historical information about the message flow **160** from the entry in the flow cache, and deletes the entry.

Flow Cache

FIG. **3** shows data structures for use with a method for routing in networks responsive to message flow patterns.

A flow cache **300** comprises a memory which associates flow keys **310** with information about message flows **160** identified by those flow keys **310**. The flow cache **300** includes a set of buckets **301**. Each bucket **301** includes a linked list of entries **302**. Each entry **302** includes information about a particular message flow **160**, including routing, access control, accounting, special treatment for packets **150** in that particular message flow **160**, and a pointer to information about treatment of packets **150** to the destination device **130** for that message flow **160**.

In a preferred embodiment, the flow cache **300** includes a relatively large number of buckets **301** (preferably about 16,384 buckets **301**), so as to minimize the number of entries **302** per bucket **301** and thus so as to minimize the number of memory accesses per entry **302**. Each bucket **301** comprises a four-byte pointer to a linked list of entries **302**. The linked list preferably includes only about one or two entries **302** at the most.

In a preferred embodiment, each entry **302** includes a set of routing information, a set of access control information, a set of special treatment information, and a set of accounting information, for packets **150** in the message flow **160**.

The routing information comprises the output port for routing packets **150** in the message flow **160**.

The access control information comprises whether access is permitted for packets **150** in the message flow **160**.

The accounting information comprises a time stamp for the first packet **150** in the message flow **160**, a time stamp for the most recent packet **150** in the message flow **160**, a cumulative count for the number of packets **150** in the message flow **160**, and a cumulative count for the number of bytes **150** in the message flow **160**.

IP Address Cache

FIG. **4** shows an IP address cache for use with a method for routing in networks responsive to message flow patterns.

An IP address cache **400** comprises a tree having a root node **410**, a plurality of inferior nodes **410**, and a plurality of leaf data structures **420**.

Each node **410** comprises a node/leaf indicator **411** and an array **412** of pointers **413**.

The node/leaf indicator **411** indicates whether the node **410** is a node **410** or a leaf data structure **420**; for nodes **410** it is set to a "node" value, while for leaf data structures **420** it is set to a "leaf" value.

The array **412** has room for exactly 256 pointers **413**; thus, the IP address cache **400** comprises an M-trie with a branching width of 256 at each level. M-tries are known in the art of tree structures. IP addresses comprise four bytes, each having eight bits and therefore 256 possible values. Thus, each possible IP address can be stored in the IP address cache **400** using at most four pointers **413**.

The inventors have discovered that IP addresses in actual use are unexpectedly clustered, so that the size of the IP address cache **400** is substantially less, by a factor of about five to a factor of about ten, than would be expected for a set of randomly generated four-byte IP addresses.

Each pointer **413** represents a subtree of the IP address cache **400** for its particular location in the array **412**. Thus, for the root node **410**, the pointer **413** at location **3** represents IP addresses having the form 3.xxx.xxx.xxx, where "xxx" represents any possible value from zero to 255. Similarly, in a subtree for IP addresses having the form 3.xxx.xxx.xxx, the pointer **413** at location **141** represents IP addresses having the form 3.141.xxx.xxx. Similarly, in a subtree for IP addresses having the form 3.141.xxx.xxx, the pointer **413** at location **59** represents IP addresses having the form 3.141.59.xxx. Similarly, in a subtree for IP addresses having the form 3.141.59.xxx, the pointer **413** at location **26** represents the IP address 3.141.59.26.

Each pointer **413** is either null, to indicate that there are no IP addresses for the indicated subtree, or points to an inferior node **410** or leaf data structure **420**. A least significant bit of each pointer **413** is reserved to indicate the type of the pointed-to structure; that is, whether the pointed-to structure is a node **410** or a leaf data structure **420**. In a preferred embodiment where pointers **413** must identify an address which is aligned on a four-byte boundary, the two least significant bits of each pointer **413** are unused for addressing, and reserving the least significant bit for this purpose does not reduce the scope of the pointer **413**.

Each leaf data structure comprises information about the IP address, stored in the IP address cache **400**. In a preferred embodiment this information includes the proper processing for packets **150** addressed to that IP address, such as a determination of a destination port for routing those packets and a determination of whether access control permits routing those packets to their indicated destination.

Flow Data Export

FIG. **5** shows a method for collecting and reporting information about message flow patterns.

A method **500** for collecting and reporting information about message flow patterns is performed by the routing device **140**.

At a flow point **510**, the routing device **140** is disposed for obtaining information about a message flow **160**. For example, in a preferred embodiment, as noted herein, the routing device **140** obtains historical information about a message flow **160** in the step **242**. In alternative embodiments, the routing device **140** may obtain information about message flows **160**, either in addition or instead, by occasional review of entries in the flow cache, or by directly monitoring packets **150** in message flows **160**.

It will be clear to those skilled in the art, after perusing this application, that the concept of reporting information about message flows is quite broad, and encompasses a wide variety of possible alternatives within the scope and spirit of the invention. For example, in alternative embodiments, information about message flows may include bi-directional traffic information instead of unidirectional traffic information, information about message flows may include information at a different protocol layer level other than that of transport service access points and other than that at which the message flow is itself defined, or information about message flows may include actual data transmitted as part of the message flow itself. These actual data may include one or more of the following: information in packet headers, information about files of file names transmitted during the message flow, or usage conditions of the message flow (such as whether the message flow involves steady or bursty transmission of data, or is relatively interactive or relatively unidirectional).

At a step **521**, the routing device **140** obtains historical information about a particular message flow **160**, and records that information in a flow data table.

At a step **522**, the routing device **140** determines a size of the flow data table, and compares that size with a selected size value. If the flow data table exceeds the selected size value, the routing device **140** continues with the step **523** to report flow data. If the flow data table does not exceed the selected size value, the routing device **140** returns to the step **521** to obtain historical information about a next particular message flow **160**.

At a step **523**, the routing device **140** builds an information packet, responsive to the information about message flows **160** which is recorded in the flow data table.

At a step **524**, the routing device **140** transmits the information packet to a selected destination device **130** on the network **100**. In a preferred embodiment, the selected destination device **130** is determined by an operating parameter of the routing device **140**. This operating parameter is set when the routing device **140** is initially configured, and may be altered by an operator of the routing device **140**.

In a preferred embodiment, the selected destination device **130** receives the information packet and builds (or updates) a database in the format for the RMON protocol. The RMON protocol is known in the art of network monitoring.

At a flow point **530**, a reporting device **540** on the network **100** is disposed for reporting using information about message flows **160**.

At a step **531**, the reporting device **540** queries the selected destination device **130** for information about message flows **160**. In a preferred embodiment, the reporting device **540** uses the RMON protocol to query the selected destination device **130** and to obtain information about message flows **160**.

At a step **532**, the reporting device **540** builds a report about a condition of the network **100**, responsive to information about message flows **160**.

At a step **533**, the reporting device **540** displays or transmits that report about the condition of the network **100** to interested parties.

In preferred embodiments, the report may comprise one or more of a wide variety of information, and interested parties

may use that information for one or more of a wide variety of purposes. Some possible purposes are noted herein:

Interested parties may diagnose actual or potential network problems. For example, the report may comprise information about packets **150** in particular message flows **160**, including a time stamp for a first packet **150** and a time stamp for a last packet **150** in the message flow **160**, a cumulative total number of bytes in the message flow **160**, a cumulative total number of packets **150** in the message flow **160**, or other information relevant to diagnosing actual or potential network problems.

Interested parties may determine patterns of usage of the network by date and time or by location. For example, the report may comprise information about which users or which services on the network are making relatively heavy use of resources. In a preferred embodiment, usage of the network **100** is displayed in a graphical form which shows use of the network **100** in a false-color map, so that network administrators and other interested parties may rapidly determine which services, which users, and which communication links are relatively loaded or relatively unloaded with demand.

Interested parties may determine which services are accessed by particular users, or which users access particular services. For example, the report may comprise information about which services are accessed by particular users at a particular device on the network **100**, or which users access a particular service at a particular device on the network **100**. This information may be used to market or otherwise enhance these services. In a preferred embodiment, users who access a particular world wide web page using the HTTP protocol are recorded, and information is sent to those users about changes to that web page and about further services available from the producers of that web page. Providers of the particular web page may also collect information about access to their web page in response to date and time of access, and location of accessing user.

Information about patterns of usage of the network, or about which services are accessed by particular users, or which users access particular services, may be used to implement accounting or billing for resources, or to set limits for resource usage, such as by particular users, by particular service providers, or by particular protocol types (and therefore by particular types of services).

Interested parties may determine usage which falls within (or without) selected parameters. These selected parameters may involve access during particular dates or times, such as for example access to particular services during or outside normal working hours. For example, it may be desirable to record those accesses to a company database which occur outside normal working hours.

These selected parameters may involve access to prohibited services, excessive access to particular services, or excessive use of network resources, such as for example access to particular servers using the HTTP protocol or the FTP protocol which fall within (or without) a particular administrative policy. For example, it may be desirable to record accesses to repositories of games or other recreational material, particularly those accesses which occur within normal working hours.

These selected parameters may involve or lack of proper access, such as for example access control list failures or unauthorized attempts to access secure services. For example, it may be desirable to record unauthorized attempts to access secure services, particularly those attempts which form a pattern which might indicate a concerted attempt to gain unauthorized access.

In alternative embodiments, the routing device **140** could save the actual packets **150** for the message flow **160**, or some part thereof, for later examination. For example, a TELNET session (a message flow **160** comprising use of the TELNET protocol by a user and a host) could be recorded in its entirety, or some portion thereof, for later examination, e.g., to diagnose problems noted with the network or with the particular host.

In further alternative embodiments, the routing device **140** could save the actual packets **150** for selected message flows **160** which meet certain selected parameters, such as repeated unauthorized attempts to gain access.

In embodiments where actual packets **150** of the message flow **160** are saved, it would be desirable to perform a name translation (such as a reverse DNS lookup), because the IP addresses for the source device **120** and the destination device **130** are transitory. Thus, it would be preferable to determine the symbolic names for the source device **120** and the destination device **130** from the IP addresses, so that the recorded data would have greater meaning at a later time.

Alternative Embodiments

Although preferred embodiments are disclosed herein, many variations are possible which remain within the concept, scope, and spirit of the invention, and these variations would become clear to those skilled in the art after perusal of this application.

We claim:

1. A method for routing messages in a data network wherein a set of packets is isolated for specialized policy treatment by a plurality of routing devices in the data network, the method comprising the steps of:

identifying a first one message of a first plurality of messages associated with an application layer, said first plurality of messages having at least one policy treatment in common, said first plurality of messages being identified in response to an address of a selected source device and an address of a selected destination device, wherein said policy treatment comprises at least one of the access control information, security information, queuing information, accounting information, traffic profiling information, and policy information;

generating a unique hash key by each of the routing devices that receives the first plurality of messages, the unique hash key being based upon the address of the selected source device, the address of the selected destination device, a port number associated with the selected source device, a port number associated with the selected destination device, and a protocol type corresponding to the first plurality of messages;

recording said first policy treatment by building a corresponding entry in a flow cache, wherein the first plurality of messages is identified by the unique hash key;

recording information about said first plurality of messages;

transmitting said information to at least one selected device on said network based upon a predetermined operating parameter;

identifying a second one message of said first plurality of messages; and

routing said second one message responsive to said first routing treatment.

2. A method as in claim **1**, wherein

said first one message comprises a packet;

said first plurality of messages comprises a stream of packets associated with a selected source device and a selected destination device.

3. A method as in claim 2, wherein said stream of packets is associated with a first selected port number at said source device and a second selected port number at said destination device.

4. A method as in claim 1, wherein said first plurality of messages comprises a message flow.

5. A method as in claim 1, wherein said first plurality of messages comprises an ordered sequence, and said first one message has a selected position in said ordered sequence.

6. A method as in claim 1, wherein said step of recording comprises building an entry flow cache, wherein said flow cache includes a plurality of entries, one said entry for each said plurality of messages, each said entry including a unicast destination address.

7. A method as in claim 1, including a step of identifying a first packet of a second stream of packets, wherein the packets of said second stream of packets have at least one second policy treatment in common, said second routing treatment differing from said first policy treatment.

8. A method as in claim 1, wherein said policy treatment comprises a destination output port for routing said first message.

9. A method as in claim 1, wherein said information comprises

an arrival time for an initial one message in said plurality of messages;

an arrival time for most recent one message in said plurality of messages;

a cumulative count of bytes in said plurality of messages; or

a cumulative count of said one messages in said plurality of messages.

10. A method as in claim 1, comprising the steps of

receiving said information at said selected device on said network;

recording said information in a database at said selected device; and

making said information available to a second device on said network.

11. A system for routing packets in a data network wherein a set of packets is isolated for specialized policy treatment, said system comprising:

a source device for outputting a stream of packets;

a destination device for receiving said stream of packets; and

a plurality of routing devices for transporting said stream of packets from said source device to said destination device, each of said plurality of routing devices comprising,

means for receiving said stream of packets, said stream of packets comprising a plurality of message flows associated with an application layer, each said packet being associated with one selected message flow, each said message flow having at least one policy treatment in common, wherein said policy treatment comprises at least one of access control information, security information, queuing information, accounting information, traffic profiling information, and policy information;

means for associating packets with a first one of said message flows,

means for generating a unique hash key upon receipt of the stream of packets, the unique hash key being based upon an address of the source device, an address of the destination device, a port number associated with the source device, a port number associated with the destination device, and a protocol type corresponding to the first plurality of messages,

means for caching an entry associated with said first one of said message flows, wherein said first one of said message flows is identified by the unique hash key,

means for recording information about said first one of said message flows;

means for transmitting said information to the destination device on said network based upon a predetermined operating parameter, and

means for routing packets responsive to entries in said caching means.

12. A system as in claim 11, wherein said entry comprises access control information.

13. A system as in claim 12, wherein said entry comprises a destination output port for routing packets.

14. A system as in claim 11, wherein said information comprises

a transmission time for an initial one message in said plurality of messages;

a transmission time for a most recent one message in said plurality of messages;

a cumulative count of bytes in said plurality of messages; or

a cumulative count of said one messages in said plurality of messages.

15. The system as in claim 11,

wherein the caching means comprises a plurality of buckets, each bucket including a linked list that includes a maximum of two entries.

16. A method for routing messages in a data network wherein a set of packets is isolated for specialized policy treatment by plurality devices in the data network, said method comprising the steps of:

identifying a first one packet of a first stream of packets defining a first message flow associated with an application layer, wherein said first stream of packets comprise an ordered sequence and said first packet has a selected position in said ordered sequence, said first stream of packets having at least one first routing policy treatment in common, wherein said policy treatment comprises at least one of access control information, security information, queuing information, accounting information, traffic profiling information, and policy information; and

generating a unique hash key by each of the routing devices that receives the first stream of packets, the unique hash key being based upon an address of a selected source device, an address of a selected destination device, a port number associated with the selected source device, a port number associated with the selected destination device, and a selected protocol type, said first routing treatment being identified by the unique hash key;

recording said unique hash key by building an entry in a flow cache;

identifying subsequent packets of a said first stream of packets defining said first message flow;

recording information about said first stream of packets;

transmitting said information to at least one selected device on said network based upon a predetermined operating parameter; and

routing said subsequent packets responsive to said first policy treatment.

17. A method as in claim 16, comprising the step of identifying a first one packet of a second stream of packets defining a second message flow, said second stream of packets having at least one second policy treatment in common, said second policy treatment differing from said first policy treatment.

18. A method as in claim 16, wherein said policy treatment further comprises a destination output port for routing said first one packet.

19. A method as in claim 16, wherein said information comprises

a transmission time for said first packet of said first stream of packets;

a transmission time for a most recent one packet in said first stream of packets;

a cumulative count of bytes in said first stream of packets; or

a cumulative count of packets in said first stream of packets.

* * * * *